



Technical Document: An Updated Methodology for Distributing Personal Income

Marina Gindelsky

DistributionofPersonalIncome@bea.gov

The distribution of personal income is a topic of widespread interest, including for policymakers, economists, and statisticians. Though literature on the topic has dramatically increased over the past two decades, led particularly by Piketty and Saez (2003), income inequality has long been a topic of interest (Kuznets 1941, 1953). Indeed, the predecessor of the Bureau of Economic Analysis (BEA), the Office of Business Economics, published estimates of the “Size Distribution of Income” covering select years between the mid-1940s and the 1970s (Office of Business Economics 1953).

Recently, interest has particularly grown in the distributional national accounts. As each household member has some claim on the economic resources of the household and affects decisions regarding economic activities, the household is regarded by the Census Bureau and in the System of National Accounts (2008) as the institutional unit for compiling distributional results (Organisation for Economic Co-operation and Development 2018). When considering households as a unit of measurement and national accounts as the aggregate, personal income represents

a natural and appropriate income concept for decision-making (Fixler, Gindelsky, and Johnson 2019). Personal income is the income received by, or on behalf of, all persons from all sources—from participation as laborers in production, from owning a home or business, from the ownership of financial assets, and from government and business in the form of transfers. It includes income from domestic sources as well as from the rest of the world; it does not include realized or unrealized capital gains or losses.

There are many metrics that can be used to measure the distribution of personal income and quantify income inequality. They are summarized in data files on the [BEA website](#) for each year from 2000 to 2019 in tables 1 and 2, in addition to in a summary table of inequality metrics for all years (table 3). Table 1, “Major Components of Personal Income and Disposable Personal Income by Decile,” distributes the household income portion of National Income and Product Accounts (NIPA) table 2.9, “Personal Income and Its Disposition by Households and by Nonprofit Institutions Serving Households,” by decile in two ways. First, households are ranked by equivalized (household size-adjusted) personal income and assigned to deciles (the first 10 numerical columns). Next, households are re-ranked by equivalized disposable personal income (personal income less taxes) and assigned to deciles. Table 1 also includes the totals for each line item in the first column as well as summary lines for personal income and disposable personal income.¹

By providing shares of income by both personal income decile and disposable personal income decile, we are able to glean additional insights from the income distribution, both as compared to the commonly provided quintiles (Census Bureau tables A3 and A4) and relative to solely pretax income. Table 2, “Inequality Metrics,” reports inequality statistics for household income, personal income, and disposable personal income.¹ A subset of summary metrics is available for all years

¹ Note that table 2 was table 3 in the March 2020 and December 2020 releases. The original “table 2” is no longer part of the

in table 3.

This is a broad set of statistics that allows for a complete picture of the distribution of income, including the share in each quintile, the top 1 percent share, the top 5 percent share, the Gini index, and the ratio of the 90th percentile to the 10th percentile.² These statistics are featured in the literature (including Piketty, Saez, and Zucman 2018; Atkinson, Piketty, and Saez 2011; Auten and Splinter 2019) as well as the World Inequality Database, Federal Reserve Board accounts (Distributional Financial Accounts), Census Bureau (aforementioned tables A3 and A4) and Bureau of Labor Statistics (BLS) publications (Cunningham 2015).

Below is a description of the methodology used to estimate the distribution of personal income, as found in the NIPA tables, specifically table 2.1 (“Personal Income and Its Disposition”), line 1. Section 1 describes the overall strategy, including how external datasets are used to calculate the statistics. Sections 2 through 7 outline the major components of personal income. Section 8 describes methodological updates to the previous release, and section 9 summarizes the data sources.

1. Overall Strategy

The Annual Social and Economic Supplement of the Current Population Survey (a base microdata set, hereafter referred to as “CPS”) was used for the analysis. There are many national accounts components that must be allocated to households (that is, households are given an amount for each component). These components can be categorized in four broad categories, “Decomposition of Personal Income for Households,” on the [BEA website](#): (1) adjusted money income (AMI), (2) financial (F), (3) health (H), and (4) other transfers (net) (T). Some of these components were allocated to households using additional data in conjunction with CPS variables. In order to maximize transparency and utility to data users, all

release. The first 12 rows are for total (unequalized) income and the final 3 rows are for equalized income.

² An appendix table with inequality statistics calculated based on equalized income is available upon request.

data used were public-use data and may be accessed using the links provided in section 9, “Data Sources.”

The CPS, jointly sponsored by the Census Bureau and BLS, is a nationally representative annual household survey of the civilian noninstitutionalized population (approximately 92,000 households sampled, representing 129 million households for 2019) that includes many demographic and income questions and is administered in February, March, and April of each year. According to the [Census Bureau](#), the survey is the primary source of labor force statistics for the United States.² It is important to note the income data in the CPS is 1 year behind the data collection year (that is, the survey year). Our analysis starts in 2000 due to the availability of our component data sources, particularly Medicare data (see section 9).³

Overall, the estimation strategy comprises the following four-steps: (1) identify a NIPA total to be distributed, (2) identify CPS variable(s) and external variables that could be used to allocate this total, (3) sum all component NIPA totals to personal income and subtotals of interest, and (4) construct inequality statistics. This process yields a dataset in which each CPS household has a value for each component of personal income. After all components have been added together to compute personal income, equivalized personal income is calculated by dividing personal income by the square root of the number of household members. For example, if household income is \$10,000 and there are four members of the household, equivalized household income is \$5,000 (half of \$10,000). By equivalizing, we are able to arrive at comparable figures for all households (that is, it is an adjustment for household size). Equivalized rankings of income are used for all income inequality metrics.

³ During the period of analysis here, there has been a significant redesign. We incorporated the first portion of the redesign in 2014 and the second in 2018. This redesign significantly affects the results, raising inequality.

2. Allocations Based on External Datasets

Several additional datasets are used to allocate NIPA totals to households.

A. Adjustment for very high incomes

Data from the Internal Revenue Service (IRS) Statistics of Income (SOI) program is used to adjust the top (that is, highest incomes) of the income distribution for three main reasons. First, the CPS is believed to unsuccessfully survey those with very high incomes, which introduces nonresponse bias in inequality estimates (Bollinger et al. 2018). Second, there is a perception that there is underreporting by those with top incomes. Third, the CPS has top codes, which vary by year, for those with top incomes so as not to risk identification of those individuals. For example, if an individual reports an income of \$10 million annually, they may be assigned a value of \$1 million by Census. For these reasons, it becomes prudent to adjust CPS incomes (Armour et al. 2021). The adjustment process is described below and is different from previous exercises (Fixler et al. 2017; Fixler, Gindelsky, and Johnson 2018, 2019, 2020; Gindelsky 2020a, 2020b).

Prior to beginning the SOI aggregation procedure outlined in Gindelsky (2020b, section 2A), we first account for misreporting present in the reported SOI data. As discussed in detail by DeBacker et al. (2020), Johns and Slemrod (2010), Auten and Splinter (2019), and others, audit data from the IRS National Research Program (NRP) and Taxpayer Compliance Measurement Program have shown that administrative data is significantly underreported (that is, the existence of a “tax gap”) due to tax noncompliance. In fact, BEA does make an adjustment for aggregate misreporting of proprietors’ income and partnership income in the macrodata, seen in NIPA table 7.14, line 2. However, the extent of the misreporting differs substantially by income source (for example, self-employment versus wage income) and varies significantly across the distribution.⁴

⁴ DeBacker et al. (2020) provide estimates (2006–2014) of misreporting by income category for wages, dividends, Schedule C, Schedule D, Schedule E, and “All Other” in Table A4, “Percentage Changes in Income as a Result of Audit by Income Source and AGI Group.”

Thus, we first multiply the SOI data for wages, self-employment, dividends (and interest by the same distribution), and rents and royalties by each adjustment factor to be consistent with the audited totals.⁵

Next, we aggregate the SOI data on tax units (see section 7) into the proportion of income with adjusted gross income (AGI) less than \$200,000, \$200,000–\$500,000, and AGI greater than or equal to \$500,000 for six components:⁶ (1) wages, (2) self-employment income,⁷ (3) ordinary dividends, (4) taxable and nontaxable interest income, (5) farm income, and (6) rents and royalties. These shares sum to 1 for each income category in each tax year. For example, a 2019 CPS that asks respondents about 2018 income would correspond to tax year 2018 SOI data. For each of the six components listed above (simply referred to as Source X below), the procedure is as follows:

⁵ Although we are able to roughly adjust the SOI data to reflect audited totals for these categories, we are not able to do a detailed adjustment of the SOI data to reflect the higher “true” totals that would include underreported income that examiners were unable to detect (Johns and Slemrod 2010). IRS researchers employ an econometric technique called “detection-controlled estimation” (DCE) to model the true amounts. BEA received a report from IRS reflecting this “true” misreporting for proprietors’ income and partnership income, which is used to adjust business income in NIPA table 7.14. However, though Johns and Slemrod (2010) do provide some distributional estimates of the misreporting impact (reported versus “true”) in their tables, they do so for percentiles and for tax year 2001. Thus, we are unable to use these for our current exercise. In the future, we hope to have these types of detailed distributional estimates for the “true” data for multiple income sources in the current period.

⁶ We run TAXSIM (a National Bureau of Economic Research (NBER) program which calculates tax liability from survey data) for all the years to impute an AGI based on original CPS values for wages, self-employment, and others. This imputed AGI is very highly correlated with the CPS AGI variable for the years after 2010 due to improvements in Census modelling. We do not have information on capital gains in the CPS for 2008–2017, which is used in TAXSIM to calculate AGI, thus our AGI would be slightly understated for those years.

⁷ Here, by self-employment, we mean both self-employment income present on IRS Schedule C and partnership income present on IRS Schedule K. Though it is not possible to distinguish between these sources in the self-employment income question in the CPS, we believe survey respondents include both types of income. To be consistent with BEA macro totals, we are subtracting the net loss column from the net income column in the SOI data for the totals used in the next step for self-employment, farm, and rents and royalties.

1. Create tax units⁸ in the CPS and sort by AGI into three groups:
 - AGI greater than or equal to \$500,000⁹
 - AGI of \$200,000–\$500,000
 - AGI less than \$200,000
2. Sum Source X (weighted) in the CPS and subtract from the NIPA total.
3. A portion (the percent SOI share for Source X) of this aggregate difference is then allocated to each income group proportional to its share of Source X in the CPS total.
4. The original CPS values I added to this “extra” imputed value.

For example, in 2012, 52 percent of dividends in the reported SOI data are for tax units with at least \$500,000 of AGI. The relevant NIPA total in this case is the SOI total of \$260 billion,¹⁰ while total weighted CPS dividend income is \$123 billion. That leaves \$260 billion minus \$123 billion = \$137 billion to be allocated to CPS households as follows: 52 percent of \$137 billion = \$71 billion to households with incomes that are at least \$500,000, 13.5 percent of \$137 billion = \$19 billion to households with incomes that are \$200,000–\$500,000, and 34.4 percent of \$137 billion = \$47 billion to households with incomes less than \$200,000. Each household then receives extra dividend income proportional to its share of dividend income

⁸ As defined by the [Urban Institute](#), “A tax unit is an individual, or a married couple, that files a tax return or would file a tax return if their income were high enough, along with all dependents of that individual or married couple. A tax unit is therefore different than a family or a household in certain situations. For example, a cohabiting couple constitutes one household but if the individuals are not legally married, they would file separate tax returns and thus be considered two tax units. A family could consist of a married couple and the wife's elderly mother who lives with them. That family would be considered two tax units since, if the elderly mother had a large enough income, she would be required to file a federal income tax return on her own. In general, the number of tax units tends to be larger than the number of families or households reported elsewhere.”

⁹ Households are sorted into tax units, using information on household structure and filing status, and AGI is recomputed using TAXSIM. If the proportion of tax units with AGI of at least \$500,000 is slightly too low to represent the proportion in the SOI tables, a subset of high earnings households is iteratively added to the group until this proportion is roughly equivalent to that of the SOI for the given year. We do the same with AGI of \$200,000–\$500,000. We do not do this for households with lower incomes, given that it may well be that households at the lower end do not file taxes but still report earned incomes to the CPS. Our approach thus becomes a weighted average of the CPS and SOI data. Tax units are re-aggregated into households afterwards such that when aggregated across all households for *Source X*, the total will add up to the NIPA total.

¹⁰ Total personal dividends are distributed in three separate parts. The portion of the total for personal dividends tie to (1) SOI-reported dividends—distributed according to the CPS dividends variable, (2) imputed dividends—distributed the same as imputed interest, and (3) residual dividends—as these are mainly tied to S-corporation shareholders, they are distributed according to the proprietors’ income distribution. The example above refers to (1), i.e., it does not include imputed dividends.

in its group such that aggregate weighted household dividend income (original income plus extra income) will sum to \$137 billion.

B. Adjustment for underreporting of certain government programs

In the current exercise, we model the reciprocity of Supplemental Security Income (SSI), Medicaid, and Supplemental Nutrition Assistance Program (SNAP) benefits to each individual in the CPS for a given survey year. From administrative data on enrollments of these programs available from the [Social Security Administration](#), [Kaiser Family Foundation](#), and [U.S. Department of Agriculture](#), respectively, we first identify the gap in underreporting. We then use the general approach of the [Congressional Budget Office \(CBO\)](#) (Habib 2018) to determine who in the CPS is most likely to underreport program participation, while using a more modern machine-learning method to improve the identification of those individuals. We use a Random Forest algorithm, which can model varied nonlinearities in the data. Program participation probabilities are made using out-of-sample predictions to remove biases from overfitting. Finally, we impute program income for these individuals, such that it corresponds to state-level NIPA totals.

In earlier versions of our exercise (Gindelsky 2020a, 2020b), we used a crosswalk, provided directly by [CBO](#), that uses an algorithm to assign probabilities of receipt of SSI, Medicaid, and SNAP to each individual in the CPS for a given survey year and then correspondingly imputes values. However, this crosswalk was unavailable for 2019 at the time of estimate construction and is not published on a regular schedule. Given the importance of the transfers in household income and documented underreporting in the CPS, we modeled the benefits to match as closely as possible. Our resulting overall reciprocity and program income means are 99 percent correlated with those of CBO for 2000–2018. On a household level, program income reciprocity and means are correlated in the microdata at more than 70 percent for all programs. However, we have greater household-level

variance in our imputations due to our process of controlling to state-level NIPA program totals, available in NIPA table SAINC35.

C. Imputation of financial income

We use the Summary Extract Public Dataset of the Survey of Consumer Finances (SCF), as obtained from the Federal Reserve Board of Governors Economic Research [website](#).¹¹ We use this dataset to distribute the three imputed components of personal income, which will be described for each relevant component. Because the SCF is triennial, we first interpolate the SCF variables for the years in which the SCF is not observed.¹² We match the SCF survey year to the CPS survey year (both income variables are lagged 1 year, while asset variables are contemporaneous). First, households are binned by income (not including capital gains). These eight bins are constructed analogously in the CPS. In order to allocate imputed interest received from banks, credit agencies, and investment companies, we first calculate the total amount of “banking” (checking plus + savings balance) and “insurance” (cash value of life insurance). Then, the relative share of each income bin to total income is imputed for each bin and matched to households in the CPS by corresponding income bin. Thus, the NIPA totals for imputed interest for these two categories are then allocated proportionally by income.

In order to allocate imputed interest and dividends from pensions, we first calculate the shares of retirees with defined contribution (DC) and defined benefit (DB) plans, respectively. Next, we calculate the share of DC participants who also have DB plans and the share of disbursements that comes from DC plans. Prior to the CPS survey redesign, the relevant information is (1) whether an individual is participating in a retirement plan or (2) whether an individual receives

¹¹ Additionally, we use imputations of defined benefit pensions available [here](#) in an SCF crosswalk, provided by the Fed

¹² We use the Fernandez procedure in the R package “tempdisagg.” The Fernandez procedure extends the Denton and Chow-Lin approaches by obtaining its solution by minimizing a quadratic loss function in the differences between the series to be created and a linear combination of the high-frequency series. The results of this method very closely match the results of the interpolation used in the Federal Reserve Board Distributional Financial Accounts.

disbursements from a retirement plan. It is not reported in the CPS whether an individual is participating in a DC or DB plan explicitly. Thus, individuals reporting participation in a retirement plan are randomly allocated to DC alone, DB alone, or DB and DC, until the share of those participating in these plans matches that of the SCF. Once individuals have been allocated to DC, DB, or both, they are given a share of the imputed interest and dividends. Similarly, a portion of those reporting “Regular payments from IRA, V .KEOGH, or 401(k) accounts” is assigned to the group of those receiving disbursements from DC plans, consistent with the share in the SCF. It is possible to reasonably identify those receiving disbursements from DB plans in the CPS in that same question. Recipients are given a share of the imputed interest and dividends according to their reported disbursement amount.

After the CPS survey redesign (implemented in 2018), it is possible to identify those contributing to retirement plans as well as those receiving disbursements from DC plans. It is, however, not possible to identify those participating in both DB and DC plans, and thus an imputation is made here as well, similar to the above strategy. Participants in DB plans are allocated a share of imputed interest and dividends according to their wage, while participants in DC plans are allocated a share according to their reported contribution. Those receiving disbursements are allocated a share according to their reported disbursement.¹³

D. Imputation of rental income for owner-occupied housing

We use the Consumer Expenditure Survey (CE) from BLS to impute rental income for owner-occupied housing for 2000–2004. Using this data source, we first rank “consumer units” (roughly the same as households) by before-tax income, creating deciles. We next construct a share of rental equivalence to before-tax family income. For example, if a household’s income is \$100,000 and it reports the expected rental value of its home is \$4,000 monthly (\$48,000 annually), the

¹³ For 2019 estimates, imputed interest and dividends for retirees were allocated along with monetary interest and dividends (rather than separately) due to differences in interest and dividend reporting in the CPS.

rent-to-income share would be 48 percent. The median share is calculated for each income decile. This share is then applied to income deciles (ranked by money income) in the CPS for households that own their home to impute a value of rental income for owner-occupied housing based on their income.

We use the American Community Survey (ACS) from Census in a similar way as the CE to impute rental income for owner-occupied housing for 2005–2019. The ACS is preferable to the CE for this calculation, given the sample size of the survey. Though the ACS does not contain a rental equivalence question, we use the new BEA strategy of Rassier et al. (2021) to impute a value for each household. We use stacked 5-year Public Use Microdata Sample (PUMS) files to have the largest possible sample size. ACS households are ranked on money income (as closely as possible to the CPS definition) and then sorted into before-tax deciles. Then, the share of rental equivalence is calculated as above for the CE. We note that there was no available 5-year PUMS file at the time of the latest release to provide estimates for 2019, and thus 2019 rental equivalence shares for imputation are kept consistent with those of 2018.

E. Employer contribution to health insurance and Medicare

In the latest CPS redesign, the Census Bureau no longer provides an imputation for the employer contribution to health insurance. Accordingly, we impute this variable ourselves (from 2010 forward). We use the MEPSnet Query Tool in order to create tables of average premium per enrolled employee and percent of total premiums contributed by employees by family size (single or family), state, year, and firm size at private sector establishments. If CPS respondents report that the employer pays the full premium, the average employer premium is imputed by state, year, family size, and firm size for the respondent. If CPS respondents report that the employer pays part of the premium and report the amount they have paid out of pocket, the employer contribution is then imputed using the average percent of the premium that employees pay by state, year, family size, and firm

size for the respondent.

Additionally, we use MEPS microdata in order to calculate the average individual Medicare expenditure by year, age category, and disability status. This is then imputed to each CPS individual reporting Medicare reciprocity, prior to being scaled to match the NIPA totals (by state). Although these nonpublic data are provided to BEA by Centers for Medicare & Medicaid Services (CMS), we have compiled a state-level crosswalk, available upon request. We have switched to this data source in order to extend the series back to 2000 and create a more nuanced imputation. The previously used publicly available CMS average expenditure data files were unavailable before 2007.

As noted above, the following sections will describe the core aggregated income components. A numeric breakout by relative size is available for the December 2020 release in the [working paper](#).

3. Adjusted Money Income (AMI)

In these estimates, we adjust money income (as defined by Census) in order to be consistent with the concepts used in the NIPA estimates. It is helpful to compare the definitions of Census money income and personal income:

“Census money income is defined as income received on a regular basis (exclusive of certain money receipts such as capital gains) before payments for personal income taxes, social security, union dues, Medicare deductions, etc. Therefore, money income does not reflect the fact that some families receive part of their income in the form of noncash benefits, such as food stamps, health benefits, subsidized housing, and goods produced and consumed on the farm.” (From “About [Income & Poverty](#)” on the Census website)

“Personal income is the income that persons receive in return for their provision of labor,

land, and capital used in current production, plus current transfer receipts less contributions for government social insurance (domestic). Personal income is equal to national income minus corporate profits with IVA and CCA_{adj}, taxes on production and imports less subsidies, contributions for government social insurance, net interest and miscellaneous payments on assets, business current transfer payments (net), and current surplus of government enterprises, plus personal income receipts on assets and personal current transfer receipts.” (From [Chapter 2 of Concepts and Methods of the U.S. National Income and Product Accounts](#))

Though Census money income, in many ways, is a narrower definition of income, it does include variables that are not in personal income, such as retirement disbursements. Accordingly, we add up the components of Census money income that are in personal income, excluding variables such as retirement disbursements (*ret-val*) or certain sources of disability income (*dis-val*) and survivor income (*sur-val*).¹⁴ We call this approximation adjusted money income.¹⁵ It is primarily comprised of income from wages and salaries, self-employment (farm and nonfarm), interest, dividends, and social security income, which together sum to 95 percent of adjusted money income in 2018. The remaining 5 percent is comprised of income from additional sources such as rents and royalties, unemployment insurance, and disability income, among others.

We distribute the components that make up AMI individually. Wages and salaries, farm income, nonfarm income, rental income of persons (other private business), interest income, and dividend income are distributed by the relevant variables, with the SOI adjustment referred to in section 2A. Federal benefits, including social security, unemployment insurance, railroad retirement, black lung benefits,

¹⁴ The exclusion of retirement disbursements constitutes approximately 75 percent of the money income excluded.

¹⁵ Adjusted money income also excludes potential sources of intrasectoral transfers, which would net out in the sector and are not associated with current period production, such as other financial assistance (*fin-val*), other income (*oi-val*), alimony (*alm-val*), child support (*csp-val*), and other noninstitutional educational assistance (*ed-val*). We do include incomes from these sources tied to railroad retirement, other retirement, workers’ compensation, black lung benefits, and state and local government disability.

pension benefit guaranty, veterans' benefits, and workers' compensation are distributed by the relevant variables in the CPS. State and local benefits including temporary disability insurance, public assistance, employment and training, education, and others are distributed to CPS households by their relevant variables, respectively. Household current transfer receipts from nonprofit institutions are distributed to CPS households by the value of their educational assistance in the CPS.

4. Financial Items

This section describes the methods used to impute financial items in the NIPAs to CPS households. This category is the sum of allocations for pensions and profit sharing, life insurance, rental income from owner-occupied housing, and imputed interest and dividends.

Employer contributions to pension plans and group life insurance are distributed to those who participate in employer-sponsored pension plans and health insurance, respectively, by wage. Imputed interest and dividends are distributed using an imputation derived from the SCF for banking, insurance, and retirement assets for each relevant item, as referred to in section 2C. Owner-occupied housing rental income is distributed using an imputation derived from the Consumer Expenditure Survey (2000–2004) and the American Community Survey (2005–2019) by income bracket referred to in section 2D.

5. Health Items

This section describes the methods used to impute health items in the NIPAs to CPS households. This category is the sum of allocations for employer contributions for health insurance, Medicare,¹⁶ Medicaid, military medical insurance, and other medical care payment assistance.

¹⁶ The totals for social security, Medicare, and Medicaid have been adjusted slightly down for the proportion of likely nursing home residents, not surveyed by CPS, but receiving these benefits. The residual is then distributed equally to all CPS individuals.

Employer contributions to health insurance are distributed by the corresponding CPS variable until 2010. After 2010, employer contributions to health insurance are distributed using data from MEPS, as described in section 2E. The exception is military medical insurance, which is distributed to active military members.

As our objective is to distribute the components of personal income, we must distribute the expenditure on Medicare and Medicaid to participants. Medicare is distributed by assigning average individual expenditures by year, age category, and disability status using MEPS data, as described in section 2E. Medicaid is distributed using a machine-learning approach to assign reciprocity and NIPA state-level totals to impute values, as described in section 2B. Medical assistance is distributed to those who are assigned SNAP and report Women, Infants, and Children (WIC) benefits or other assistance, except the Children's Health Insurance Program (CHIP), which provides assistance to households reporting children covered by CHIP.

6. Net Other Transfers

This section describes the methods used to impute items related to transfers in the NIPAs that are not contained in adjusted money income or health to CPS households. These items include (net) employer and employee contributions for government social insurance,¹⁷ SNAP, WIC, refundable tax credits, energy assistance, educational assistance, and other transfers.

Employer/employee contributions to old-age, survivors, disability, and hospital insurance are distributed by the imputed value of FICA (Federal Insurance Contributions Act) as calculated by TAXSIM, the NBER program for calculating tax liability from survey data. Employer contributions to workers' compensation and supplemental unemployment are distributed respectively by wage. Other employer and employee/self-employed contributions to government social

¹⁷ In table 2.1, these items are included as part of compensation (line 2) and subsequently deducted in line 25.

insurance are also distributed by wage. Military medical insurance (federal benefits and employee/self-employed contributions to government social insurance) are distributed to CPS households with active military members. SNAP is distributed using the machine-learning approach described in section 2B. Refundable tax credits are distributed to CPS households separately by the criteria for the relevant credit where possible. For example, the values (as calculated by TAXSIM) for the earned income tax credit (EITC), child tax credit (CTC), and additional child tax credit benefits are used, as is participation in the marketplace for the health premium tax credit. Together, these categories represent 95 percent of the refundable tax credits in 2018. One complication is that the TAXSIM calculations impute the credits for the previous calendar year, regardless of when they are received by the households. Thus, there can be an issue of timing for some credits and households.

Energy assistance and WIC are distributed to CPS households by the relevant indicators. Household current transfer receipts from business (net) are distributed equally to all households. All other transfers are distributed to households reporting receipt of WIC, SNAP, or other assistance.

To calculate personal income from household income, household current transfer receipts from nonprofits and nonprofit institution transfer receipts from households are deducted, and nonprofit institution income is added. This residual is distributed equally to all individuals in the CPS.

7. Taxes

The calculation of tax liability (and associated components) was done using the NBER TAXSIM program (Feenberg and Coutts 1993). CPS households were organized into tax units, whose income inputs were scaled to NIPA totals. Federal taxes before credits (*fiitax*), state taxes before credits (*siitax*), FICA, EITC, CTC, and alternative minimum tax liability were merged from TAXSIM back into the CPS for the calculations to proceed. In the latest CPS redesign, the Census imputation for

property taxes is no longer available. Thus, we impute property taxes for TAXSIM by calculating the property tax rate using the ratio of the tax variable to the property value variable in the 2014–2018 5-year ACS by public use microdata area (PUMA). Then, we use a PUMA-county crosswalk to assign the property tax rate to each county. Where a local rate is not available, due to disclosure restrictions, we impute a state-level property tax rate and assign it to those observations. Disposable personal income is calculated by subtracting federal taxes and state taxes from personal income. As a 5-year file is not yet available for 2019, the 2018 values are used.

8. Methodological Updates

There have been several methodological updates since the previous prototype release (December 2020). These methodological changes reflect expert feedback, including from participants at the NBER conference, BEA Advisory Committee meetings, and Committee on National Statistics expert meeting.

1. The strategy for allocating additional income to high-income households has been revised. As noted above in section 2A, there is now an additional AGI grouping (\$200,000–\$500,000). Thus, households are now assigned to either AGI of at least \$500,000, AGI \$200,000–\$500,000, or AGI of less than \$200,000 such that the proportion of tax units in the CPS matches those of the SOI data for those categories.
2. The totals for social security, Medicare, and Medicaid have been adjusted slightly down for the proportion of likely nursing home residents, not surveyed by CPS, but receiving these benefits. The residual is then distributed equally to all CPS households.
3. The strategy for Medicare allocation has now been modified, as described in section 2E. It is now based on averages by state, age, and disability status as calculated from microdata from MEPS. This change was made to create a more disaggregated distribution of Medicare, as well as to extend the series

back to 2000.

4. The former strategy of using CBO crosswalks to impute reciprocity, and subsequently values, for Medicaid, SNAP, and SSI, has now been replaced with a more modern machine-learning method. We impute program income for these individuals, such that it corresponds to state-level NIPA totals. Overall program participation and income means are 99 percent correlated with CBO imputations for 2000–2018, despite having a larger variance. This method is described in section 2B.
5. The strategy for allocation of imputed interest and dividends has been modified. As described in detail in section 2C, imputed interest and dividends are now allocated separately for DC and DB plans, and for retirees and nonretirees, using information in the SCF. Additionally, the number of income bins used for allocating banking and insurance has been reduced from 11 to 8 to reduce year-to-year volatility.
6. While the strategy for imputing rental equivalence has not changed, the 5-year ACS PUMS files have replaced the CE for imputations for years 2005–2019, as described in section 2D. The ACS has a significantly larger sample size and is also part of the new BEA housing estimation procedure.
7. The weighting procedure has been modified such that each decile (or quintile) contains the same number of weighted, rather than sample, households in a given CPS year.

9. Data Sources

There were many data sources (listed below with links) used for the compilation of these estimates. All are publicly available and made available with varying time lags after the data is collected. Information on incomes from the CPS and Consumer Expenditure Survey is regularly available with a 1-year time lag (for example, 2019 values became available in fall of 2020) since the surveys ask

respondents about income received in the previous calendar year. Key data from the IRS SOI program, which is used to adjust top incomes, is regularly available with a 2-year lag (for example, 2018 values became available in 2020). The SCF is a triennial survey and asks about income received in the previous year, similar to the CPS and the Consumer Expenditure Survey. While the previous versions of this exercise used a CBO crosswalk to adjust CPS values for certain transfer programs in order to correct for underreporting in these categories, there is no regular publication schedule for these estimates, and the most recent year available was 2018 at the time of estimate construction. Therefore, BEA has developed an alternative methodology for imputing receipt and expenditures. This is described in section 2B. Medical Expenditure Panel Survey (MEPS) published estimates are available with a 1-year lag, while the microdata is available with a 2-year lag.

1. **National Income and Product Accounts.** The primary source for our national macrodata for 2000 to 2019 is the BEA NIPA data on the [BEA website](#). Please note this data is subject to frequent revision. For the current set of estimates, we used the October 2021 release.
2. **Current Population Survey, March Supplement.** This is the core data series that is the base of the analysis. We used all households (except those residing in group quarters) from the survey years 2001 to 2020. This data was downloaded from the [NBER website](#) for 2000 to 2017. However, for 2018–2020, we have downloaded the CPS files directly from the Census website.
3. **Consumer Expenditure Survey.** We used this survey in order to estimate rental equivalence for owner-occupied housing from 2000 to 2014. The data was downloaded from [BLS](#).
4. **Survey of Consumer Finances.** We used this survey in order to distribute imputed interest income from the NIPAs for 2007, 2010, 2013, 2016, and 2019. For the

nonsurvey years, we imputed values with the methodology described in section 2C. The data was downloaded from the [Federal Reserve Board](#).

5. **Internal Revenue Service Statistics of Income data.** In order to adjust top incomes in the CPS, we used summary IRS Form 1040 data aggregated by AGI for tax years 2000 to 2019 in table 1.4 found in “SOI Tax Stats” on the [IRS website](#).
6. **American Community Survey 2005–2019 (stacked 5-year PUMS files).** In order to impute property tax rates for the 2018 CPS, we use the 5-year 2014–2018 public use ACS household file, as downloaded from the [Census Bureau](#). In order to calculate rental equivalence, we use the 5-year ACS files of earlier years.
7. **Medical Expenditure Panel Survey data.** Tables II.C.1, II.C.3, II.D.1, and II.D.3 were downloaded for each year 2010–2019 using the [MEPSnet Query](#) for the insurance component to impute the employer premium and percent employee contribution. MEPS microdata provided by CMS was used for Medicare expenditure. A state-level crosswalk is available upon request.

References

- Armour, Philip, Gerald Auten, Richard Burkhauser, and Jeff Larrimore. 2021. "Recent Trends in U.S. Top Income Shares in Tax Record Data Using More Comprehensive Measures of Income Including Accrued Capital Gains." *Journal of Political Economy* 129 (5): 1319–1360.
- Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49 (1): 3–71.
- Auten, Gerald, and David Splinter. 2019. "Top 1 Percent Income Shares: Comparing Estimates Using Tax Data." *AEA Papers and Proceedings* 109 (1): 307–311.
- Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak. 2019. "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch." *Journal of Political Economy* 127 (5).
- Cunningham, J. Chris. 2015. "Measuring wage inequality within and across U.S. metropolitan areas, 2003–13." *Monthly Labor Review* 138.
- DeBacker, Jason, Bradley Heim, Anh Tran, and Alexander Yuskavage. 2020. "Tax Noncompliance and Measures of Income Inequality." *Tax Notes* 166 (7): 1103–1118.
- Feenberg, Daniel, and Elizabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management* 12 (1): 189–194.
- Fixler, Dennis, David Johnson, Kevin Furlong, and Andrew Craig. 2017. "A Consistent Data Series to Evaluate Growth and Inequality in the National Accounts." *Review of Income and Wealth* 63 (S2): S437–S459.
- Fixler, Dennis, Marina Gindelsky, and David Johnson. 2020. "Measuring Inequality in the National Accounts." BEA Working Papers, WP 2020–03.
- Fixler, Dennis, Marina Gindelsky, and David Johnson. 2019. "Improving the Measure of the Distribution of Personal Income." *AEA Papers and Proceedings* 109: 302–306.
- Fixler, Dennis, Marina Gindelsky, and David Johnson. 2018. "Towards a Distribution of Household Income: Linking Survey Data to Administrative Data." Paper presented at the International Association for Research in Income and Wealth 35th General Conference, Copenhagen, Denmark, August 20–25.
- Gindelsky, Marina. 2020a. "Technical Document: A Methodology for Distributing

Personal Income.” BEA website.

Gindelsky, Marina. 2020b. “Technical Document: An Updated Methodology for Distributing Personal Income.” BEA website.

Johns, Andrew and Joel Slemrod. 2010. “The Distribution of Income Tax Noncompliance.” *National Tax Journal* 63 (3): 397–418.

Kuznets, Simon. 1941. *National Income and Its Composition, 1919–1938, Volume I*. New York: National Bureau of Economic Research.

Kuznets, Simon. 1953. *Shares of Upper Income Groups in Income and Savings*. New York: National Bureau of Economic Research.

Organisation for Economic Co-operation and Development. 2018. *Handbook on compiling distributional results on household income, consumption and saving consistent with national accounts*.

Piketty, Thomas, and Emmanuel Saez. 2003. “Income Inequality in the United States, 1913–1998.” *Quarterly Journal of Economics* 118 (1).

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. “Distributional National Accounts: Methods and Estimates for the United States.” *Quarterly Journal of Economics* 13 (2).

Rassier, Dylan G., Bettina H. Aten, Eric B. Figureoa, Solomon Kublashvili, Brian J. Smith, and Jack York. 2021. “Improved Measures of Housing Services for the U.S. Economic Account.” *Survey of Current Business* 101 (5).

U.S. Office of Business Economics. 1953. *Income Distribution in the United States, by Size, 1944– 1950*. Washington, DC: U.S. Government Printing Office.